



Transkribus

Der Einsatz von maschinellem Lernen und Handwritten Text Recognition in der Erschließung historischer Dokumente

MARC ROTHBALLER | MÜNCHEN

Während die Texterkennung von modernen Druckschriften bereits ausgereift und weitverbreitet ist, befindet sich die Erkennung von Handschriften noch in den Kinderschuhen. Mit Transkribus gibt es eine KI-basierte Anwendung, die Experten- wie auch Laiennutzern vielfältige Versprechen macht und Arbeitsabläufe vereinfachen möchte. Die Einsatzbereiche reichen von A wie Archiv über H wie Heimatforschung bis V wie Vermessung – ein Überblick und Erfahrungsbericht.

ALTE SCHRIFT UND NEUE TECHNOLOGIE

Die Vorzüge von Optical Character Recognition (OCR) für gescannte Dokumente sind aus dem Büro- und Berufsalltag bekannt: Texterkennung erschließt Dokumente, sie werden durchsuchbar und der Recherche- und Arbeitsaufwand lässt sich enorm reduzieren.

Für Antiqua-Schriften (wie z. B. Times New Roman) gibt es bereits exzellente Softwarelösungen, deren Einsatzfelder von der Smartphone-App bis hin zur Business-Lösung reichen. Bei sogenannten gebrochenen Schriften (Fraktur), deren Buchstabenbögen nicht rund, sondern spitz und von Richtungswechseln geprägt sind, sieht es anders aus. Wer ein Buch oder eine Zeitung »in alter deutscher Schrift« texterkennen lassen möchte, wird schnell feststellen: Hier beißen die allermeisten (kostenlosen oder kostengünstigen) OCR-Programme auf Granit.

Dies liegt zum einen daran, dass sich Buchstabenformen teilweise gleichen: Das **ſ** (s) und das **f** (f), das **k** (k) und **t** (t) oder das **B** (B) und **V** (V) sehen sich täuschend ähnlich und ihre Lesbarkeit ist stark von der Druckqualität und Alterung abhängig. Zum anderen scheint eine Fraktur-OCR kommerziell nicht zu lohnen, denn Fraktur wird heute nicht mehr verwendet und fand auch historisch nur regional begrenzt Verbreitung. Es ist also ein Nischenmarkt, der sich auf Expertenanwender wie Bibliotheken beschränkt, die ihre frakturgesetzten Buch- und Zeitungsbestände digitalisieren möchten. Sollen keine Druckschriften, sondern Handschriften erkannt werden, muss aktuell noch jedes gängige Texterkennungsprogramm kapitulieren, spätestens bei historischen Handschriften fehlen brauchbare Lösungen beinahe gänzlich.

Dies ist gerade in Arbeits- und Anwendungsfeldern ärgerlich, in denen auf historische Dokumente zurückgegriffen werden muss, die von heutigen Nutzern grundsätzlich nicht mehr gelesen werden können. Die Lesekompetenz für alte Handschriften muss mühsam erlernt werden, ihr Fehlen erschwert das Verständnis und den Arbeitsablauf, ja verunmöglicht mitunter die Weiterarbeit. Man denke an Personenstandsunterlagen der Standesämter, die bis in die Mitte des 20. Jahrhunderts handschriftlich und in Kurrent (»alter deutscher Schreibschrift«) geschrieben wurden.

Aber auch Vermessungsunterlagen und Kataster, Baupläne, Notariats- und Vertragsdokumente sind bis ca. 1945 regelmäßig noch in Kurrent- oder Sütterlinhandschrift verfasst und damit für nach 1945 geborene Personen nicht mehr lesbar. Wie gut, dass mit Transkribus eine Softwarelösung zur Verfügung steht, die es einem abnimmt, selbst noch einmal zum ABC-Schützen zu werden – oder?

TRANSKRIBUS – WUNDERKIND DER HANDSCHRIFTENERKENNUNG?

Transkribus ist eine Anwendung bzw. vielmehr eine teilweise cloudbasierte Plattform zur (halb) automatisierten Layout- und Texterkennung von Archivunterlagen und historischen Dokumenten mittels maschinenlernender Verfahren. Es wurde von der Universität Innsbruck als Teil des EU-geförderten READ-Projekts bis Juni 2019 entwickelt und wird nun von der Europäischen Genossenschaft READ-COOP SCE betrieben. Neben der reinen Handschriften- und Texterkennung bietet Transkribus noch weitere Funktionen (Metadatenanreicherung, Keyword Spotting, kollaboratives Arbeiten usw.), die im Folgenden jedoch vernachlässigt werden.

Transkribus verspricht eine Lücke zu füllen, die oben erwähnte Anwendungen bisher offen gelassen haben und die unterschiedlichen Personenkreisen (Historikern, beruflichen Anwendern, Heimat- und Familienforschern, Citizen Scientists) bisher die Arbeit erschwert – es ist daher naheliegend, dass Transkribus mit dem Wolpertinger als »Wappentier« wirbt.

Transkribus erscheint als vielversprechender Hoffnungsträger, der Einsatz sogenannter künstlicher Intelligenz (KI) bzw. maschinellen Lernens verheißt zudem stetiges Hinzulernen und verbesserte (autonome) Leistung. Der ursprünglich kostenlose Zugang zu Transkribus – Archive und Bibliotheken müssen bereits seit längerem für die Nutzung zahlen – wurde im Oktober 2020 auch für Privatnutzer auf ein Bezahlmodell umgestellt.

Doch so wie sich ein Wolpertinger nach einer bekannten Jagdregel nur an abgelegenen Waldesrändern durch eine junge und gut aussehende Frau sichten lässt, die sich bei Vollmond in der Abenddämmerung mit einem statthaften Mann auf die Pirsch begibt, so hat auch der Wolpertinger Transkribus seine Tücken: Wer sich eine einfach zu bedienende Anwendung à la Adobe Acrobat Pro erhofft, mit der sich »mal schnell« ein paar gescannte oder abfotografierte Dokumente ohne Mehraufwand transkribieren lassen, wird enttäuscht werden. Ebenfalls enttäuscht werden Anwender ohne Kompetenzen im Lesen alter Schriften.

Die Anwendung kann einiges an Texterkennung automatisch erledigen, ganz so, wie es beispielsweise die OCR von Adobe Acrobat, ABBYY FineReader oder vergleichbaren Programmen mit Antiqua-Schriften macht. Zur Überprüfung und Vervollständigung des erkannten handschriftlichen Textes muss man aber erkennen können, wo Transkribus falschlag und was stattdessen wirklich geschrieben steht.

Auch eine gewisse Technikaffinität und Geübtheit mit Java-Anwendungen schaden bei der Benutzung von Transkribus nicht. Leider ist die Anwendung optisch wenig ansprechend gestaltet:

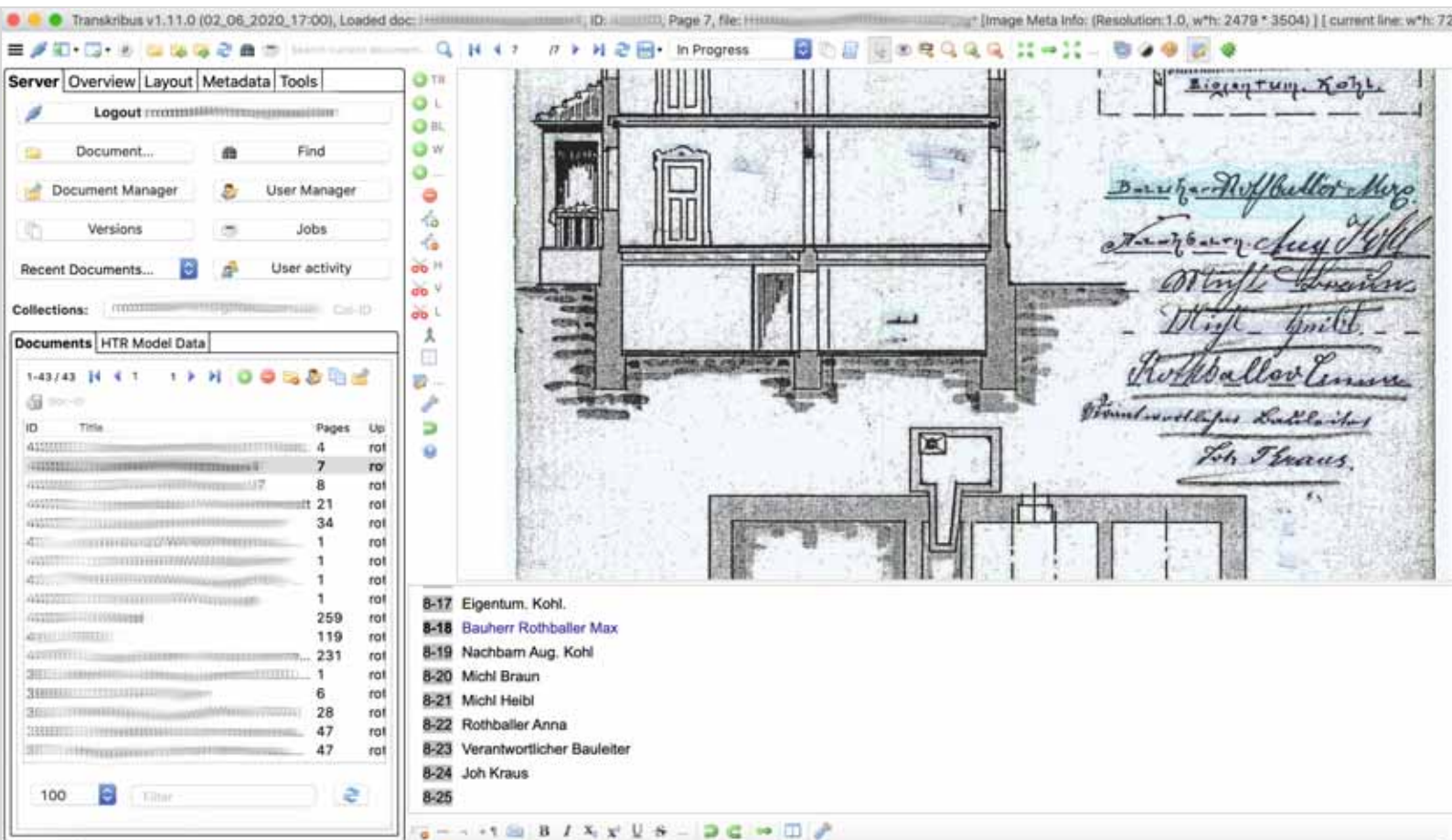


Abbildung 1 | Grafische Benutzeroberfläche (GUI) von Transkribus mit Bauplan (1910) und Transkription

Sie ist überfrachtet mit Symbolen und allerlei Bedienelementen, die Menüanordnung, die Bedienung selbst und die Durchführung der Arbeitsabläufe sind reichlich unintuitiv, umständlich und nicht selbsterklärend.

Aufgrund der Java-Programmierung und der serverbasierten Auftragsverarbeitung erledigt die Anwendung ihre Arbeit zudem recht gemächlich und es wird für beinahe jeden Schritt eine Internetverbindung benötigt – zugleich ist sie dadurch aber weitgehend plattformunabhängig und es können auch umfangreiche Dokumente erkannt werden, ohne damit die eigene Rechnerleistung über Stunden zu belasten. Ein Wiki sowie Tutorials führen in die Bedienung ein.

Wenngleich sich die Anwendung an Experten (Historiker, Archive etc.) richtet, so erscheint sie auch für diese Zielgruppe unnötig unübersichtlich und verstaubt. Gewiss, über Geschmack lässt sich streiten – Transkribus wirkt auf den ersten Blick allerdings eher abschreckend als einladend und es bleibt zu hoffen, dass zukünftige Versionen durch eine grundlegende optische Überarbeitung und Verschlanung »Usability statt Frustration« (Lehnenmeier, Burghardt 2019) verheißßen. Dennoch sind seine Vorzüge schon jetzt nicht von der Hand zu weisen.

WAS MACHT TRANSKRIBUS?

Nach Installation und Registrierung können Dokumente hochgeladen werden. Die Dokumente (PDF, JPEG, TIFF u. a.) werden auf den Transkribus-Servern gespeichert/verarbeitet und sind nur für den Nutzer sichtbar, laut Anbieter werden die europäischen Datenschutzrichtlinien erfüllt.

Im weiteren Prozess kommen maschinelle Verfahren zum Einsatz, die datengetrieben arbeiten. Maschinelles Lernen ist eine grundlegende Methode der KI. Dabei soll eine Maschine automatisiert sinnvolle Ergebnisse zu einer Problemstellung liefern, ohne dass ihr zuvor der konkrete Lösungsweg aufgezeigt wurde. Dies funktioniert, indem ein Algorithmus – eine Folge von Anweisungen – aus vorliegenden Beispieldaten ein Modell erlernt, das sich dann auf neue und zuvor noch nicht gesehene Daten anwenden lässt.

Je mehr Daten initial zur Verfügung stehen, desto besser das Endergebnis. Die Daten, die Transkribus benötigt, sind segmentierte Zeilen und eine korrekte Transkription. Vereinfacht ausgedrückt erlernt Transkribus, dass eine bestimmte Form, Gestaltung und Folge von Linien mit einem bestimmten Druckbuch-

staben gleichzusetzen ist. Es eignet sich ein Muster, eine »Buchstaben-Blaupause« an, anhand derer es weiter trainiert und auf die richtige Fährte geführt wird. Statt von »künstlicher Intelligenz« wäre es hier angebracht von »maschinellern Lernen« durch eine Mustererkennungsmaschine zu sprechen.

Intelligenz würde nämlich auch einschließen, dass die Maschine den Text versteht, den sie transkribiert, dass ihr auffällt, wenn im gescannten Dokument eine Seite fehlt und der Übergang von einer zur nächsten Seite keinerlei Sinn ergibt, ja womöglich sogar, dass sie von den Zeilen und Wörtern, die sie da umwandelt, wahlweise ergriffen, angewidert, gelangweilt oder gefeselt wäre – und vieles mehr. Das tut Transkribus aber nicht. Es ärgert sich auch nicht, wenn es etwas nicht »richtig« entziffern kann, und knobelt dann lange an diesem Wort – es setzt das in Text um, was seinem erlernten Muster am nächsten kommt, und überlässt die Bewertung (und damit das Ärgern und die Nacharbeit) der menschlichen Intelligenz.

Um eine Handschriftenerkennung durchführen zu können, muss das Dokument daher zunächst segmentiert sein. Dazu werden die einzelnen Spalten, Absätze, Zeilen etc. auf jeder Seite markiert. Dies geschieht im Programm entweder vollkommen manuell (und umständlich) oder vollkommen automatisiert mit anschließender manueller Nacharbeit.

Sobald das Dokument automatisch segmentiert ist (dieser Prozess läuft über die Transkribus-Server und nimmt Zeit in Anspruch), erscheint es in der Seitenanzeige mit Zeilen/Absatzmarkierungen hinterlegt. Jede Zeile ist nun als solche mit einer Grundlinie markiert, auch die Worthöhen sind mit Linien gekennzeichnet – so weiß Transkribus, in welchen Bereichen mit Zeichen und Wörtern zu rechnen ist. Transkribus gibt den erkannten Text später entsprechend den markierten Zeilen aus, wer hier eine Zeile übersieht oder nicht automatisch markierte Wörter in der Nachbearbeitung vergisst zu markieren, wird diese auch nicht in der Textausgabe wiederfinden.

Wer markierte Flecken oder vermeintlich als Textblock erkannte Zeichnungen nicht demarkiert, der erhält zu Zeichenbrei verarbeitete Grafiken – die sogenannte KI ist aktuell nur so clever wie ihr Lehrmeister. Wurde der Text aber gänzlich segmentiert, so kann zur Handschriftenerkennung geschritten werden, der Handwritten Text Recognition (HTR). Es werden sowohl gedruckte (Fraktur-)Schriften als auch Handschriften jeglicher Art (arabisch, hebräisch, womöglich sogar die des Voynich-Manuskripts) erkannt.

Um ein schlagkräftiges Modell zu erhalten, verfügt der Anwender im Idealfall über ein paar Hundert Seiten des gleichen Schreibers. So kann speziell für diesen Schreiber bzw. dessen Handschrift ein Modell erstellt und trainiert werden. Man tran-

skribiert dazu die ersten Seiten selbst (empfohlen sind 5.000 bis 15.000 Wörter, dieser Artikel hat knapp 3.100 Wörter), füttert die Anwendung also mit Informationen und Beispielen dazu, welche handschriftliche Buchstabenform welchem Druckbuchstaben entspricht, und sie wird mittels ihres Algorithmus sodann alleine nach vergleichbaren Formationen und tolerablen Abweichungen davon im Rest des Dokuments suchen.

Transkribus »lernt«, die Handschrift zu lesen, indem es die Zeichen, die es im Dokument vorfindet (neue Situation), mit denen in der hinterlegten »Blaupause« (bekanntes Muster) vergleicht und dazu vom Nutzer Feedback erhält. Derartige Modelle erreichen mit einem entsprechenden Training eine Genauigkeit von über 90 % (vgl. Seaward, Kallio 2017).

Die Erstellung eines eigenen Modells ist für große Dokumentensammlungen (Tagebücher, umfangreiche Manuskripte und Akten etc.) von einer oder sehr wenigen Handschriften sinnvoll und gewinnbringend. So wurde beispielsweise eine umfangreiche Lesenotissammlung von Michel Foucault bearbeitet. Mit einem Trainingsset von 200 Seiten wurde eine Genauigkeit von ca. 85 % erreicht, mit 400 weiteren Seiten konnte diese auf ca. 92 % verbessert werden – bei einer als »herausfordernd« bewerteten Philosophenhandschrift (vgl. Massot et al. 2019).

Durch Verbesserungen im maschinellen Lernen ist laut Aussage der Transkribus-Entwickler aktuell bei Einzelschreibern von einer Fehlerquote von unter 5 % auszugehen. Jedoch entspricht dies nicht unbedingt dem gängigen Anwendungsfall: Häufig liegen eine Vielzahl von (kurzen) Dokumenten unterschiedlicher Schreiber aus unterschiedlichen Epochen vor. Jede Epoche ein anderes Schriftbild, jeder Schreiber seine eigene Handschrift – zu wenig Material, um für jedes Dokument eine eigene Trainingsdatei zu erstellen.

Glücklicherweise gibt es in Transkribus einige (von anderen Nutzern erstellte und trainierte) Modelle, auf die in solchen Fällen zurückgegriffen werden kann. Auch solche Basismodelle mit einer ausgewogenen Mischung aus mehreren Schreibern einer Epoche arbeiten nicht viel schlechter als Einzelmodelle und können recht passable Ergebnisse liefern. Eine Trainingsdatei zur Kurrentschrift (deutsch) aus dem 19. Jahrhundert wurde z. B. mit 3 Millionen Wörtern und 0,5 Millionen Zeilen trainiert (Stand: Juli 2020). Als Trainingsmaterial dienen aber Schwei-



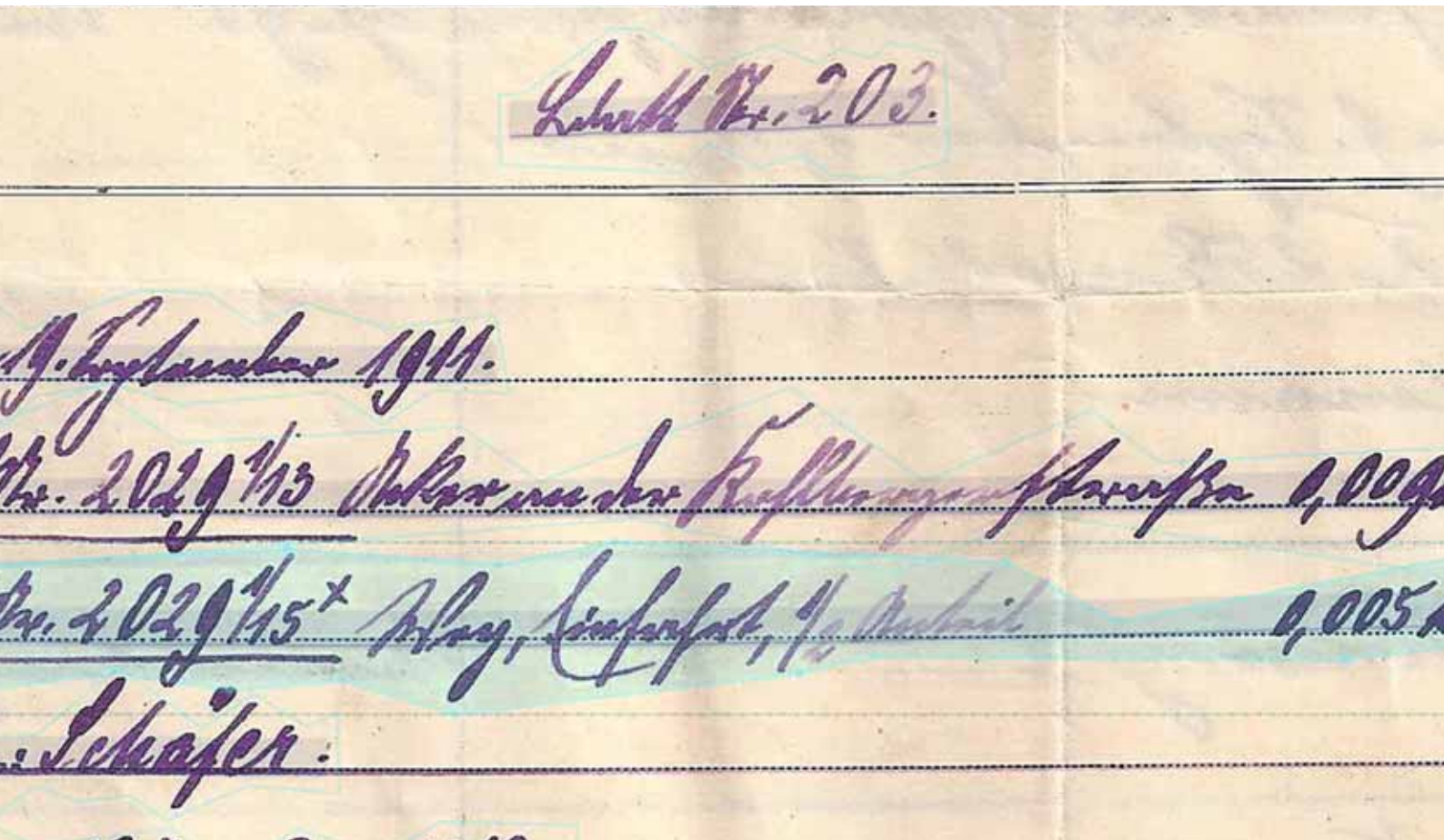


Abbildung 2 | Grundbuchauszug Etzenricht, 1911

zer Archivalien, daher hat die Trainingsdatei einen leichten »Schweiz-Bias«.

Eine weitere Trainingsdatei zur Kurrentschrift (deutsch) deckt das 16. bis 18. Jahrhundert ab, 1,5 Millionen Wörter und 250.000 Zeilen sind erfasst (Stand: Juli 2020). Das Modell »NZZ Gold Standard« wurde mit Titelseiten der »Neuen Zürcher Zeitung« von 1780–1940 trainiert und erkennt Fraktur praktisch fehlerlos. So gibt es noch eine Reihe weiterer Modelle (auch für andere Sprachen), die für eine pragmatische Grunderschließung eines Textkorpus geeignet sein können.

Letztlich benötigt Transkribus (noch) eine »Anleitung«, mit der es das vorliegende Dokument einordnen und »verstehen«, will heißen: entschlüsseln, kann. Wird für ein deutschsprachiges und kurrent geschriebenes Dokument aus dem 18. Jahrhundert vom Nutzer ein Modell gewählt, das mit schwedischen Gerichtsunterlagen des 19. Jahrhunderts trainiert wurde, dann verweigert sich Transkribus dem nicht – in der Texterkennung kommt letztlich jedoch wenig Sinnvolles heraus.

Folglich gilt es also, entweder ein eigenes Modell zu trainieren oder ein einigermaßen passendes Modell (Entstehungsort und

-sprache, Entstehungszeit, Schriftart) zu wählen und dieses über den Text laufen und seinen Dienst tun zu lassen. Dies geschieht wieder über die Transkribus-Server und dauert je nach Umfang des Dokuments einige oder viele Minuten, pro Seite kann man mit 20 bis 50 Sekunden rechnen. Nach Abschluss der Texterkennung wird passend zum abgebildeten Dokument der erkannte Text angezeigt.

Dieser Text kann sodann überarbeitet und nachgebessert werden – das sollte er auch, denn von Perfektion kann (noch) keine Rede sein. Ein nicht auf die spezifische Handschrift trainiertes Modell liefert eine Erkennungsrate von 50 bis 80 %. Freilich variiert die Trefferquote je nach Qualität der Aufnahme, Lesbarkeit und Gleichmäßigkeit der Handschrift sowie korrekter Segmentierung. Eine Nachbearbeitung ist also (noch) unabdingbar und setzt voraus, dass der Anwender prinzipiell in der Lage ist, der Anwendung auf die Sprünge zu helfen – nämlich indem er die Schrift selbst entziffert und Fehler korrigiert.

Soweit ein eigenes Modell trainiert wird, lernt Transkribus dadurch und verbessert stetig seine Leistungsfähigkeit. Aufgrund der cloudbasierten Verortung profitieren davon langfristig alle Nutzer.

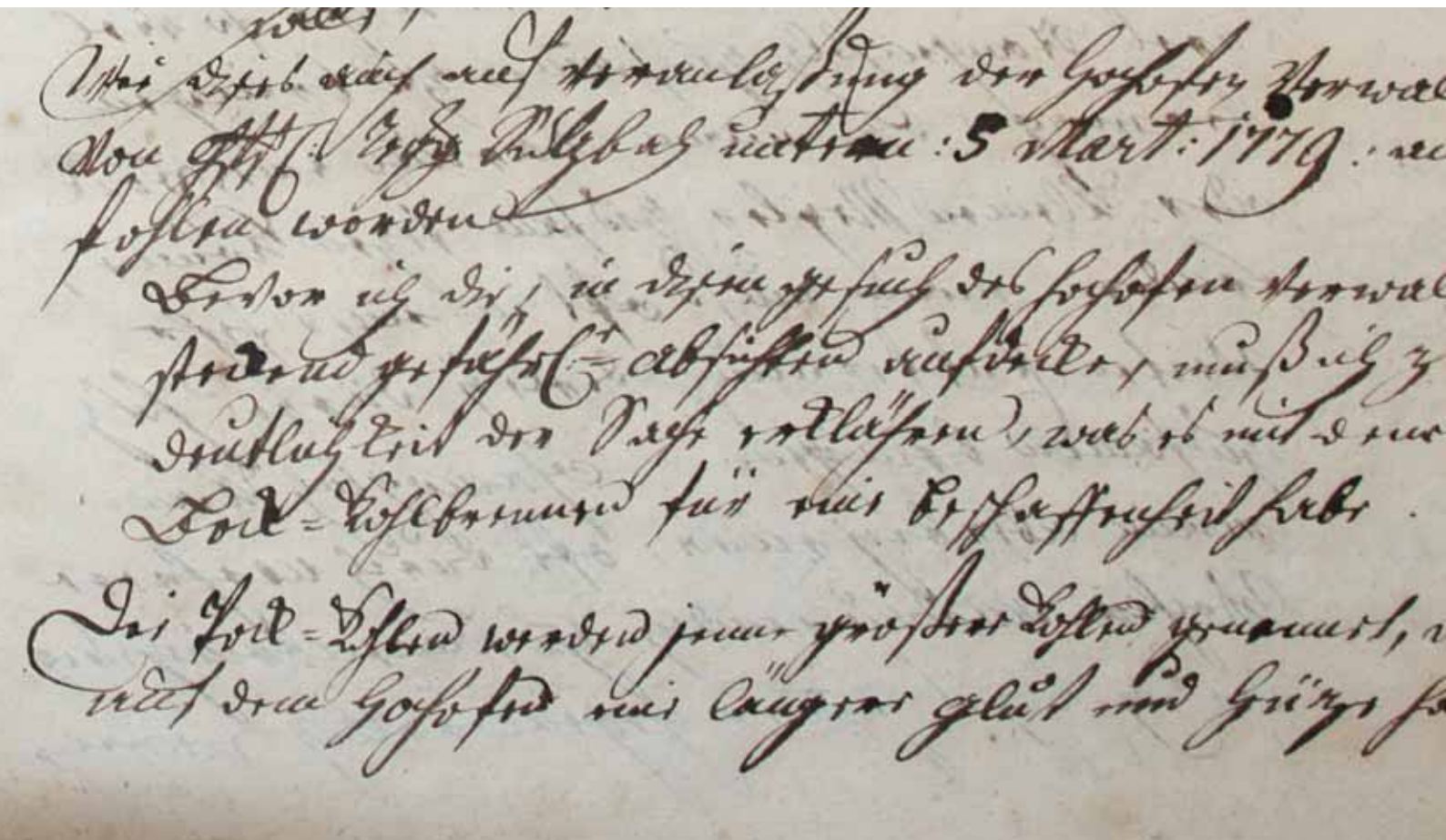


Abbildung 3 | Privatbesitz (NL-Z), Köhlerei und Meilerei,
Schreiben des Franz Benno von Hann v. 18.2.1780 an die Reg. Pfalz-Sulzbach, fol. 13r

»DEN PODL-AHLEN WERDEND SEINE GRÖSSERE BOHLEND GENANNET«

Nun aber zwei Beispiele: zunächst die Klageschrift des Revierförsters von Mantel bei Weiden in der Oberpfalz von 1780 an seinen Fürsten in Sulzbach (Abbildung 3). Der Leseschwierigkeitsgrad der Handschrift ist als »normal« zu bewerten, sie ist weder sonderlich klar oder verspielt noch kritzelig. Sie enthält Jahreszahlen, allerlei Abkürzungen und Symbole sowie fantasie-reiche Orthografie. Das Dokument ist frei von größeren Flecken oder Verunreinigungen (Schmutz, Wasserflecken, Tintenschmierer, Schimmel- oder Fehlstellen etc.), stellenweise scheint die rückseitige Schrift durch. Das Dokument wurde mit einer DSLR (24 MP, ohne Stativ) fotografiert, auch Aufnahmen mit einem High-End-Smartphone bei Tages-/Archivlicht sind problemlos verwertbar. Die entstandenen Bilddateien wurden gespeichert und in PDF konvertiert. Die Uploadmöglichkeiten (u. a. FTP, DFG Viewer METS-URL, Local) überfordern den Laiennutzer schnell, erfreuen aber den Experten.

An dem so erstellten Dokument wurde keine weitere Bildbearbeitung durchgeführt, die Aufnahmen sind in Farbe und ohne Anpassung von Kontrast etc. weiterverarbeitet worden.

Was Transkribus erkennt:

Wie ders auch auf veranlaßung der hochofen verwaltun
von Gen. Weg Sulzbach mit den 5. Mart. 1779.
anbeohlen worden
Bevor ich die, in disen gesuch des hochofen verwalters
stellend gefahren absichten aufdelle, muß ich zur
Deutlichkeit der Sache erklären, was es mit denen
Bodt-Kohlbrennend für eine beschaffenheit habe
Den Podl-ahlen werdend seine größere Bohlend genannet,
die auf dem hochofen und längere glust und Hürs halten

Was tatsächlich geschrieben steht:

Wie dises auch auf Veranlaßung der Hochofen Verwaltung
von chfstl. Regg. Sulzbach untern : 5 Mart. : 1779 :
anbefohlen worden.
Bevor ich die in disen gesuch des hochofen Verwalters
steckend gefährl. absichten aufdecke, muß ich zur
Deutlichkeit der Sache erklären, was es mit dem
Bock-Kohlbrennen für eine Beschaffenheit habe.
Dei Pock-Kohlen werden jenne größere Kohlen genannet,
die auf dem Hochofen eine längere Glut und Hütze halten



Bei dem anderen Beispiel (*Abbildung 2*) handelt es sich um einen Grundbuchauszug aus dem Jahr 1911. Das Formblatt ist mit Fraktur bedruckt und handschriftlich in Kurrent ausgefüllt. Das Dokument wurde mit 300 dpi gescannt und als PDF in Transkribus importiert.

Der Frakturdruck wird sehr zuverlässig erkannt, die Handschrift bzw. vor allem die Abkürzungen und Zahlen bereiten der Anwendung jedoch Probleme. So wird aus der Datumsangabe »1911« das Jahr »1941«, mit »PINr. 2029« weiß Transkribus nichts anzufangen und erkennt es als »Al. 201945«. Die darunterstehende »PINr. 2029« macht es gar zu »Peller. 202945«.

Bei Konsonanten und Wortendungen muss häufiger nachgearbeitet werden, abgekürzte Wörter verwirren die KI gelegentlich, mit Zahlen hat sie mal Probleme und mal erkennt sie sie erstaunlich gut. Gerade fachspezifische Abkürzungen oder Symbole (Währungszeichen, Maßeinheiten etc.) müssen trainiert werden.

Eine Nachbearbeitung ist also dringend geboten – man sollte sich aber immer vor Augen halten, dass hier eine Maschine anhand von einigermaßen vergleichbaren Trainingsdaten eine menschliche Handschrift »erkennt«, die noch dazu Wörter und Schreibweisen verwendet, die selten in Wörterbüchern zu finden sind (was eine Erkennung zusätzlich vereinfachen würde). Mit einem speziell trainierten Modell sind weitaus weniger Fehler zu erwarten.

Unter diesen Gesichtspunkten erstaunt das Ergebnis durchaus, zumal mit jedem »Training«, mit jeder Korrektur, die Schrifterkennung präziser wird. Gleichwohl wird ein »öffentliches« Modell nicht um die Lerneffekte einer individuellen Anwendung erweitert. Das fertig (nach-)bearbeitete Dokument lässt sich u. a. als durchsuchbares PDF (also als Bild mit Text hinterlegt) oder auch nur als Reintext speichern.

WAS BRINGT TRANSKRIBUS?

Für den unbedarften Laien ohne Lesekompetenz bringt die Anwendung zunächst kaum einen Mehrwert. Zumindest grundlegende Lesekompetenzen und die Bereitschaft, diese auszubauen, müssen vorhanden sein. Wer handschriftliche Dokumente transkribieren (und weiterverarbeiten) möchte, der wird

in Transkribus eine bedenkenswerte Möglichkeit sehen, Prozesse zu verkürzen und den Faktor »menschliche Ermüdung« in der Transkription zu reduzieren. Gleichwohl ist eine manuelle Nachbearbeitung (noch) unabdingbar und die mühsame Bedienung von Transkribus gestaltet diese nicht angenehm. Gerade wenn umfangreiche Materialsammlungen eines oder weniger Schreiber transkribiert werden sollen, kann durch das Training eines eigenen Modells jedoch sehr viel Zeit bei der Transkription und Nachbearbeitung gespart werden.

Ob der Einsatz von Transkribus auch wirtschaftlich ist, muss gut kalkuliert werden. Denn die Handschriftenerkennung schlägt nach Verbrauch des für 400 Seiten reichenden Freikontingents mit knapp 20 Cent je Seite zu Buche. Ausschlaggebend ist zudem (noch) der Nachbearbeitungsaufwand durch die menschliche Intelligenz.


Ohne Zweifel ist die Digitalisierung von Archivalien unter vielerlei Gesichtspunkten ein wichtiger Schritt: Dokumente werden orts- und zeitunabhängig sowie barrierefrei zugänglich gemacht, vom damit verbundenen konservatorischen Schutz der Originale ganz zu schweigen. Die Digitalisierung an sich wird aber ebenso zweifellos nur der erste Schritt hin zu einer KI-basierten (Teil-)Aufbereitung der digitalisierten Archivalien sein. Maschinelles Lernen setzt voraus, dass es zumindest initial einen Lehrer gibt, der mit genügend Lehrmaterial und Know-how bewaffnet der KI den Weg weist. Auf absehbare Zeit wird durch das anhaltende Training der Transkribus-Algorithmen und die technische Weiterentwicklung des maschinellen Lernens eine noch größere Genauigkeit bei (noch weniger) Zu- und Nacharbeit zu erzielen sein. Hiervon werden Experten, aber auch und vor allem Laienanwender profitieren.

Literatur

Marie-Laure Massot et al. (2019): *Transcribing Foucault's handwriting with Transkribus*. In: *Journal of Data Mining and Digital Humanities*, hal.archives-ouvertes.fr/hal-01913435v3/document.

Constantin Lehenmeier, Manuel Burghardt (2019): *Usability statt Frustration. Eine Fallstudie zur Usability von Digital Humanities-Tools am Beispiel der OCR-Software Transkribus*.

In: **Draude, Lange, Sick (Hrsg.):** *INFORMATIK 2019 Workshops, Lecture Notes in Informatics (LNI)*, Gesellschaft für Informatik, Bonn 2019, S. 97–106.

Louise Seaward, Maria Kallio (2017): *Transkribus. Handwritten Text Recognition technology for historical documents*, pdfs.semanticscholar.org/e9d7/e94bcc43063536e034c6d1b2f5391f3d8e42.pdf. 



Marc Rothballer
München
rothballer.marc@gmail.com